**ARTICLE**

# On mappings on the hypercube with small average stretch

Lucas Boczkowski[1] and Igor Shinkar[2,*]

[1]CNRS, IRIF Université Paris 7, France and [2]Simon Fraser University, Burnaby, BC, V5A 1S6, Canada
*Corresponding author. Email: ishinkar@sfu.ca

**Abstract**

Let $A \subseteq \{0,1\}^n$ be a set of size $2^{n-1}$, and let $\phi : \{0,1\}^{n-1} \to A$ be a bijection. We define *the average stretch* of $\phi$ as

$$\mathsf{avgStretch}(\phi) = \mathbb{E}[\mathsf{dist}(\phi(x), \phi(x'))],$$

where the expectation is taken over uniformly random $x, x' \in \{0,1\}^{n-1}$ that differ in exactly one coordinate. In this paper, we continue the line of research studying mappings on the discrete hypercube with small average stretch. We prove the following results.

- For any set $A \subseteq \{0,1\}^n$ of density $1/2$ there exists a bijection $\phi_A : \{0,1\}^{n-1} \to A$ such that $\mathsf{avgStretch}(\phi_A) = O\left(\sqrt{n}\right)$.

- For $n = 3^k$ let $A_{\mathsf{rec\text{-}maj}} = \{x \in \{0,1\}^n : \mathsf{rec\text{-}maj}(x) = 1\}$, where $\mathsf{rec\text{-}maj} : \{0,1\}^n \to \{0,1\}$ is the function *recursive majority of 3's*. There exists a bijection $\phi_{\mathsf{rec\text{-}maj}} : \{0,1\}^{n-1} \to A_{\mathsf{rec\text{-}maj}}$ such that $\mathsf{avgStretch}(\phi_{\mathsf{rec\text{-}maj}}) = O(1)$.

- Let $A_{\mathsf{tribes}} = \{x \in \{0,1\}^n : \mathsf{tribes}(x) = 1\}$. There exists a bijection $\phi_{\mathsf{tribes}} : \{0,1\}^{n-1} \to A_{\mathsf{tribes}}$ such that $\mathsf{avgStretch}(\phi_{\mathsf{tribes}}) = O(\log(n))$.

These results answer the questions raised by Benjamini, Cohen, and Shinkar (Isr. J. Math 2016).

**Keywords:** boolean hypercube; lipschitz mappings

**2020 MSC Codes:** Primary: 68R05

## 1. Introduction

In this paper, we continue the line of research from [2, 15, 9] studying geometric similarities between different subsets of the hypercube $\mathcal{H}_n = \{0,1\}^n$. Given a set $A \subseteq \mathcal{H}_n$ of size $|A| = 2^{n-1}$ and a bijection $\phi : \mathcal{H}_{n-1} \to A$, we define ***the average stretch*** of $\phi$ as

$$\mathsf{avgStretch}(\phi) = \mathbb{E}_{x \sim x' \in \mathcal{H}_{n-1}} \left[ \mathsf{dist}\left(\phi(x), \phi\left(x'\right)\right) \right],$$

where $\mathsf{dist}(x, y)$ is defined as the number of coordinates $i \in [n]$ such that $x_i \neq y_i$, and the expectation is taken over a uniformly random $x, x' \in \mathcal{H}_{n-1}$ that differ in exactly one coordinate.[1]

The origin of this notion is motivated by the study of the complexity of distributions [5, 17, 12]. In this line of research, given a distribution $\mathcal{D}$ on $\mathcal{H}_n$ the goal is to find a mapping $h : \mathcal{H}_m \to \mathcal{H}_n$

---

[1]Note that any $C$-Lipschitz function $\phi : \mathcal{H}_{n-1} \to A$ satisfies $\mathsf{avgStretch}(\phi) \leq C$. That is, the notion of average stretch is a relaxation of the Lipschitz property.

such that if $U_m$ is the uniform distribution over $\mathcal{H}_m$, then $h(U_m)$ is (close to) the distribution $\mathcal{D}$, and each output bit $h_i$ of the function $h$ is computable efficiently (e.g., computable in $AC_0$, i.e., by polynomial size circuits of constant depth).

Motivated by the goal of proving lower bounds for sampling from the uniform distribution on some set $A \subseteq \mathcal{H}_n$, Lovett and Viola [12] suggested the restricted problem of proving that no bijection from $\mathcal{H}_{n-1}$ to $A$ can be computed in $AC_0$. Toward this goal they noted that it suffices to prove that any such bijection requires large average stretch. Indeed, by the structural results of [7, 4, 11] it is known that any such mapping $\phi$ that is computable by a polynomial size circuit of depth $d$ has $\mathsf{avgStretch}(\phi) < \log{(n)}^{O(d)}$, and hence proving that any bijection requires super-polylogarithmic average stretch implies that it cannot be computed in $AC_0$. Proving a lower bound for sampling using this approach remains an open problem.

Studying this problem, [2] have shown that if $n$ is odd, and $A_{\mathsf{maj}} \subseteq \mathcal{H}_n$ is the hamming ball of density $1/2$, that is $A_{\mathsf{maj}} = \{x \in \mathcal{H}_n : \sum_i x_i > n/2\}$, then there is a $O(1)$-bi-Lipschitz mapping from $\mathcal{H}_{n-1}$ to $A_{\mathsf{maj}}$, thus suggesting that proving a lower bound for a bijection from $\mathcal{H}_{n-1}$ to $A_{\mathsf{maj}}$ requires new ideas beyond the sensitivity-based structural results of [7, 4, 11] mentioned above. In [15] it has been shown that if a subset $A_{\mathsf{rand}}$ of density $1/2$ is chosen uniformly at random, then with high probability there is a bijection $\phi : \mathcal{H}_{n-1} \to A_{\mathsf{rand}}$ with $\mathsf{avgStretch}(\phi) = O(1)$. This result has recently been improved by Johnston and Scott [9], who showed that for a random set $A_{\mathsf{rand}} \subseteq \mathcal{H}_n$ of density $1/2$ there exists a $O(1)$-Lipschitz bijection from $\mathcal{H}_{n-1}$ to $A_{\mathsf{rand}}$ with high probability.

The following problem was posed in [2], and repeated in [15, 9].

**Problem 1.1.** Exhibit a subset $A \subseteq \mathcal{H}_n$ of density $1/2$ such that any bijection $\phi : \mathcal{H}_{n-1} \to A$ has $\mathsf{avgStretch}(\phi) = \omega(1)$, or prove that no such subset exists.[2]

**Remark.** Note that it is easy to construct a set of density $1/2$ such that any bijection $\phi : \mathcal{H}_{n-1} \to A$ must have a *worst case stretch* at least $n/2$. For example, for odd $n$ consider the set $A = \{y \in \mathcal{H}_n : n/2 < \sum_i y_i < n\} \cup \{0^n\}$. Then any bijection $\phi : \mathcal{H}_{n-1} \to A$ must map some point $x \in \mathcal{H}_{n-1}$ to $0^n$, while all neighbours $x'$ of $x$ are mapped to some $\phi(x')$ with weight at least $n/2$. Hence, the *worst case stretch* of $\phi$ is at least $n/2$. In contrast, Problem 1.1 does not seem to have a non-trivial solution.

To rephrase Problem 1.1, we are interested in determining a tight upper bound on the $\mathsf{avgStretch}$ that holds uniformly for all sets $A \subseteq \mathcal{H}_n$ of density $1/2$. Note that since the diameter of $\mathcal{H}_n$ is $n$, for any set $A \subseteq \mathcal{H}_n$ of density $1/2$ and any bijection $\phi : \mathcal{H}_{n-1} \to A$ it holds that $\mathsf{avgStretch}(\phi) \le n$. It is natural to ask how tight this bound is, that is, whether there exists $A \subseteq \mathcal{H}_n$ of density $1/2$ such that any bijection $\phi : \mathcal{H}_{n-1} \to A$ requires linear average stretch.

It is consistent with our current knowledge (though hard to believe) that for any set $A$ of density $1/2$ there is a mapping $\phi : \mathcal{H}_{n-1} \to A$ with $\mathsf{avgStretch}(\phi) \le 2$. The strongest lower bound we are aware of is for the set $A_\oplus = \left\{x \in \mathcal{H}_n : \sum_i x_i \equiv 0 \pmod 2\right\}$. Note that the distance between any two points in $A_\oplus$ is at least $2$, and hence $\mathsf{avgStretch}(\phi) \ge 2$ for any mapping $\phi : \mathcal{H}_{n-1} \to A_\oplus$. Proving a lower bound strictly greater than $2$ for any set $A$ is an open problem, and prior to this work we are not aware of any sublinear upper bounds that apply uniformly to all sets.

Most of the research on metric embedding, we are aware of, focuses on worst case stretch. For a survey on metric embeddings of finite spaces see [10]. There has been a lot of research on the question of embedding into the Boolean cube. For example, see [1, 8] for work on embeddings between random subsets of the Boolean cube, and [6] for isometric embeddings of arbitrary graphs into the Boolean cube.

---

[2]*Throughout the paper, the density of a set $A \subseteq \mathcal{H}_n$ is defined as $\mu_n(A) = \frac{|A|}{2^n}$.*

### 1.1. A uniform upper bound on the average stretch

We prove a non-trivial uniform upper bound on the average stretch of a mapping $\phi : \mathcal{H}_{n-1} \to A$ that applies to all sets $A \subseteq \mathcal{H}_n$ of density $1/2$.

**Theorem 1.2.** *For any set $A \subseteq \mathcal{H}_n$ of density $\mu_n(A) = 1/2$, there exists a bijection $\phi : \mathcal{H}_{n-1} \to A$ such that* $\mathsf{avgStretch}(\phi) = O\left(\sqrt{n}\right)$.

Toward this goal we prove a stronger result bounding the average transportation distance between two arbitrary sets of density $1/2$. Specifically, we prove the following theorem.

**Theorem 1.3.** *For any two sets $A, B \subseteq \mathcal{H}_n$ of density $\mu_n(A) = \mu_n(B) = 1/2$, there exists a bijection $\phi : A \to B$ such that* $\mathbb{E}\left[\mathsf{dist}\left(x, \phi(x)\right)\right] \leq \sqrt{2n}$.

Note that Theorem 1.2 follows immediately from Theorem 1.3 by the following simple argument.

**Proposition 1.4.** *Fix a bijection $\phi : \mathcal{H}_{n-1} \to A$. Then* $\mathsf{avgStretch}(\phi) \leq 2\mathbb{E}_{x \in \mathcal{H}_{n-1}}[\mathsf{dist}$ $(x, \phi(x))] + 1$.

**Proof.** Using the triangle inequality we have

$$
\begin{aligned}
\mathsf{avgStretch}(\phi) &= \mathbb{E}_{\substack{x \in \mathcal{H}_{n-1} \\ i \in [n-1]}}[\mathsf{dist}(\phi(x), \phi(x + e_i))] \\
&\leq \mathbb{E}[\mathsf{dist}(x, \phi(x)) + \mathsf{dist}(x, x + e_i) + \mathsf{dist}(x + e_i, \phi(x + e_i))] \\
&= \mathbb{E}[\mathsf{dist}(x, \phi(x))] + 1 + \mathbb{E}[\mathsf{dist}(x + e_i, \phi(x + e_i))] \\
&= 2\mathbb{E}[\mathsf{dist}(x, \phi(x))] + 1,
\end{aligned}
$$

as required. $\qquad \square$

### 1.2. Bounds on the average stretch for specific sets

Next, we study two specific subsets of $\mathcal{H}_n$ defined by Boolean functions commonly studied in the field "Analysis of Boolean functions" [13]. Specifically, we study two monotone noise-sensitive functions: the *recursive majority of 3's*, and the *tribes* function.

It was suggested in [2] that the set of ones of these functions $A_f = f^{-1}(1)$ may be such that any mapping $\phi : \mathcal{H}_{n-1} \to A_f$ requires large $\mathsf{avgStretch}$. We show that for the recursive majority function there is a mapping $\phi_{\mathsf{rec\text{-}maj}} : \mathcal{H}_{n-1} \to \mathsf{rec\text{-}maj}^{-1}(1)$ with $\mathsf{avgStretch}(\phi_{\mathsf{rec\text{-}maj}}) = O(1)$. For the tribes function we show a mapping $\phi_{\mathsf{tribes}} : \mathcal{H}_{n-1} \to \mathsf{tribes}^{-1}(1)$ with $\mathsf{avgStretch}(\phi_{\mathsf{tribes}}) = O(\log(n))$. Below we formally define the functions, and discuss our results.

#### 1.2.1. Recursive majority of 3's

The recursive majority of 3's function is defined as follows.

**Definition 1.5.** *Let $k \in \mathbb{N}$ be a positive integer. Define the function **recursive majority of 3's** $\mathsf{rec\text{-}maj}_k : \mathcal{H}_{3^k} \to \{0, 1\}$ as follows.*

- *For $k = 1$ the function $\mathsf{rec\text{-}maj}_1$ is the majority function on the 3 input bits.*
- *For $k > 1$ the function $\mathsf{rec\text{-}maj}_k : \mathcal{H}_{3^k} \to \{0, 1\}$ is defined recursively as follows. For each $x \in \mathcal{H}_{3^k}$ write $x = x^{(1)} \circ x^{(2)} \circ x^{(3)}$, where each $x^{(r)} \in \mathcal{H}_{3^{k-1}}$ for each $r \in [3]$. Then, $\mathsf{rec\text{-}maj}_k(x) = \mathsf{maj}\left(\mathsf{rec\text{-}maj}_{k-1}\left(x^{(1)}\right), \mathsf{rec\text{-}maj}_{k-1}\left(x^{(2)}\right), \mathsf{rec\text{-}maj}_{k-1}\left(x^{(3)}\right)\right)$.*

Note that $\mathsf{rec\text{-}maj}_k(x) = 1 - \mathsf{rec\text{-}maj}_k(\mathbf{1} - x)$ for all $x \in \mathcal{H}_n$, and hence the density of the set $A_{\mathsf{rec\text{-}maj}_k} = \left\{x \in \mathcal{H}_n : \mathsf{rec\text{-}maj}_k(x) = 1\right\}$ is $\mu_n\left(A_{\mathsf{rec\text{-}maj}_k}\right) = 1/2$. We prove the following result regarding the set $A_{\mathsf{rec\text{-}maj}_k}$.

**Theorem 1.6.** *Let $k$ be a positive integer, let $n = 3^k$, and let $A_{\text{rec-maj}_k} = \{x \in \mathcal{H}_n : \text{rec-maj}_k(x) = 1\}$. There exists a mapping $\phi_{\text{rec-maj}_k} : \mathcal{H}_{n-1} \to A_{\text{rec-maj}_k}$ such that* $\text{avgStretch}\left(\phi_{\text{rec-maj}_k}\right) \leq 20$.

### 1.2.2. The tribes function
The tribes function is defined as follows.

**Definition 1.7.** *Let $s, w \in \mathbb{N}$ be two positive integers, and let $n = s \cdot w$. The function $\text{tribes} : \mathcal{H}_n \to \{0, 1\}$ is defined as a DNF consisting of $s$ disjoint clauses of width $w$.*

$$\text{tribes}\left(x_1, x_2, \ldots, x_w; \ldots; x_{(s-1)w+1} \ldots x_{sw}\right) = \bigvee_{i=1}^{s} \left(x_{(i-1)w+1} \wedge x_{(i-1)w+2} \wedge \cdots \wedge x_{iw}\right).$$

*That is, the function $\text{tribes}$ partitions $n = sw$ inputs into $s$ disjoint 'tribes' each of size $w$, and returns 1 if and only if at least one of the tribes 'votes' 1 unanimously.*

It is clear that $\mathbb{P}_{x \in \mathcal{H}_n}[\text{tribes}(x) = 1] = 1 - (1 - 2^{-w})^s$. The interesting settings of parameters $w$ and $s$ are such that the function is close to balanced, that is, this probability is close to $1/2$. Given $w \in \mathbb{N}$, let $s = s_w = \ln(2)2^w \pm \Theta(1)$ be the largest integer such that $1 - (1 - 2^{-w})^s \leq 1/2$. For such choice of the parameters we have $\mathbb{P}_{x \in \mathcal{H}_n}[\text{tribes}(x) = 1] = \frac{1}{2} - O\left(\frac{\log(n)}{n}\right)$ (see, e.g., [[13], Section 4.2]).

Consider the set $A_{\text{tribes}} = \{x \in \mathcal{H}_n : \text{tribes}(x) = 1\}$. Since the density of $A_{\text{tribes}}$ is not necessarily equal to $1/2$, we cannot talk about a bijection from $\mathcal{H}_{n-1}$ to $A_{\text{tribes}}$. In order to overcome this technical issue, let $A_{\text{tribes}}^*$ be an arbitrary superset of $A_{\text{tribes}}$ of density $1/2$. We prove that there is a mapping $\phi_{\text{tribes}}$ from $\mathcal{H}_{n-1}$ to $A_{\text{tribes}}^*$ with average stretch $\text{avgStretch}(\phi_{\text{tribes}}) = O(\log(n))$. In fact, we prove a stronger result, namely that the average transportation distance of $\phi_{\text{tribes}}$ is $O(\log(n))$.

**Theorem 1.8.** *Let $w$ be a positive integer, and let $s$ be the largest integer such that $1 - (1 - 2^{-w})^s \leq 1/2$. For $n = s \cdot w$ let $\text{tribes} : \mathcal{H}_n \to \{0, 1\}$ be defined as a DNF consisting of $s$ disjoint clauses of width $w$. Let $A_{\text{tribes}} = \{x \in \mathcal{H}_n : \text{tribes}(x) = 1\}$, and let $A_{\text{tribes}}^* \subseteq \mathcal{H}_n$ be an arbitrary superset of $A_{\text{tribes}}$ of density $\mu_n\left(A_{\text{tribes}}^*\right) = 1/2$. Then, there exists a bijection $\phi_{\text{tribes}} : \mathcal{H}_{n-1} \to A_{\text{tribes}}^*$ such that $\mathbb{E}[\text{dist}(x, \phi_{\text{tribes}}(x))] = O(\log(n))$. In particular, $\text{avgStretch}(\phi_{\text{tribes}}) = O(\log(n))$.*

### 1.3. Roadmap
The rest of the paper is organized as follows. We prove Theorem 1.3 in section 2. In section 3, we prove Theorem 1.6, and in section 4, we prove Theorem 1.8.

## 2. Proof of Theorem 1.3

We provide two different proofs of Theorem 1.3. The first proof, in subsection 2.1 shows a slightly weaker bound of $O\left(\sqrt{n \ln(n)}\right)$ on the average stretch using the Gale-Shapley result on the *stable marriage problem*. The idea of using the stable marriage problem was suggested in [2], and we implement this approach. Then, in subsection 2.2, we show the bound of $O\left(\sqrt{n}\right)$ by relating the average stretch of a mapping between two sets to known estimates on the Wasserstein distance on the hypercube.

### 2.1. Upper bound on the average transportation distance using stable marriage
Recall the Gale-Shapley theorem on the *stable marriage problem*. In the stable marriage problem we are given two sets of elements $A$ and $B$ each of size $N$. For each element $a \in A$ (resp. $b \in B$) we have a ranking of the elements of $B$ (resp. $A$) given as an bijection $rk_a : A \to [N]$ $\left(rk_b : B \to [N]\right)$

representing the preferences of each $a$ (resp. $b$). A matching (or a bijection) $\phi : A \to B$ is said to be *unstable* if there are $a, a' \in A$, and $b, b' \in B$ such that $\phi(a) = b'$, $\phi(a') = b$, but $rk_a(b) < rk_a(b')$, and $rk_b(a) < rk_b(a')$; that is, both $a$ and $b$ would prefer to be mapped to each other over their mappings given by $\phi$. We say that a matching $\phi : A \to B$ is *stable* otherwise.

**Theorem 2.1.** *(Gale-Shapley theorem) For any two sets $A$ and $B$ of equal size and any choice of rankings for each $a \in A$ and $b \in B$ there exists a stable matching $m : A \to B$.*

For the proof of Theorem 1.3 the sets $A$ and $B$ are subsets of $\mathcal{H}_n$ of density $1/2$. We define the preferences between points based on the distances between them in $\mathcal{H}_n$. That is, for each $a \in A$ we have $rk_a(b) < rk_a(b')$ if and only if $\mathsf{dist}(a, b) < \mathsf{dist}(a, b')$ with ties broken arbitrarily. Similarly, for each $b \in B$ we have $rk_b(a) < rk_b(a')$ if and only if $\mathsf{dist}(a, b) < \mathsf{dist}(a', b)$ with ties broken arbitrarily.

Let $\phi : A \to B$ be a bijection. We show that if $\mathbb{E}_{x \in A}[\mathsf{dist}(x, \phi(x))] > 3k$ for $k = \lceil \sqrt{n \ln(n)} \rceil$, then $\phi$ is not a stable matching. Consider the set

$$F := \{x \in A \mid \mathsf{dist}(x, \phi(x)) \geq k\}.$$

Note that since the diameter of $\mathcal{H}_n$ is $n$, and $\mathbb{E}_{x \in A}[\mathsf{dist}(x, \phi(x))] > 3k$, it follows that $\mu_n(F) > \frac{k}{n}$. Indeed, we have $3k < \mathbb{E}_{x \in A}[\mathsf{dist}(x, \phi(x))] \leq n \cdot \frac{\mu_n(F)}{\mu_n(A)} + k \cdot \left(1 - \frac{\mu_n(F)}{\mu_n(A)}\right) \leq n \cdot \frac{\mu_n(F)}{\mu_n(A)} + k$, and thus $\mu_n(F) > \frac{2k}{n} \cdot \mu_n(A)$. Next, we use Talagrand's concentration inequality.

**Theorem 2.2.** *([[16], Proposition 2.1.1]) Let $k \leq n$ be two positive integers, and let $F \subseteq \mathcal{H}_n$. Let $F_{\geq k} = \{x \in \mathcal{H}_n : \mathsf{dist}(x, y) \geq k \; \forall y \in F\}$ be the set of all $x \in \mathcal{H}_n$ whose distance from $F$ is at least $k$. Then $\mu_n(F_{\geq k}) \leq e^{-k^2/n}/\mu_n(F)$.*

By Theorem 2.2 we have $\mu_n(F_{\geq k}) \leq e^{-k^2/n}/\mu_n(F)$, and hence, for $k = \lceil \sqrt{n \ln(n)} \rceil$ it holds that

$$\mu_n(F_{\geq k}) \leq e^{-\ln(n)}/\mu_n(F) \leq (1/n)/(k/n) = 1/k.$$

In particular, since $\mu_n(\phi(F)) = \mu_n(F) > k/n > 1/k \geq \mu_n(F_{\geq k})$, there is some $b \in \phi(F)$ that does not belong to $F_{\geq k}$. That is, there is some $a \in F$ and $b \in \phi(F)$ such that $\mathsf{dist}(a, b) < k$. On the other hand, for $a' = \phi^{-1}(b)$, by the definition of $F$ we have $\mathsf{dist}(a, \phi(a)) \geq k$ and $\mathsf{dist}(a', b = \phi(a')) \geq k$, and hence $\phi$ is not stable, as $a$ and $b$ prefer to be mapped to each other over their current matching. Therefore, in a stable matching $\mathbb{E}_{x \in A}[\mathsf{dist}(x, \phi(x))] \leq 3 \lceil \sqrt{n \ln(n)} \rceil$, and by the Gale-Shapley theorem such a matching does, indeed, exist.

### 2.2. Proof of Theorem 1.3 using transportation theory

Next we prove Theorem 1.3, by relating our problem to a known estimate on the Wasserstein distance between two measures on the hypercube. Recall that the $\ell_1$-Wasserstein distance between two measures $\mu$ and $\nu$ on $\mathcal{H}_n$ is defined as

$$W_1(\mu, \nu) = \inf_q \sum_{x,y} \mathsf{dist}(x, y) q(x, y),$$

where the infimum is taken over all couplings $q$ of $\mu$ and $\nu$, that is, $\sum_{y'} q(x, y') = \mu(x)$ and $\sum_{x'} q(x', y) = \nu(y)$ for all $x, y \in \mathcal{H}_n$. That is, we consider an optimal coupling $q$ of $\mu$ and $\nu$ minimizing $\mathbb{E}_{(x,y) \sim q}[\mathsf{dist}(x, y)]$, the expected distance between $x$ and $y$, where $x$ is distributed according to $\mu$ and $y$ is distributed according to $\nu$.

We prove the theorem using the following two claims.

**Claim 2.3.** *Let $\mu_A$ and $\mu_B$ be uniform measures over the sets $A$ and $B$ respectively. Then, there exists a bijection $\phi$ from $A$ to $B$ such that $\mathbb{E}[\mathsf{dist}(x, \phi(x))] = W_1(\mu_A, \mu_B)$.*

**Claim 2.4.** *Let $\mu_A$ and $\mu_B$ be uniform measures over the sets $A$ and $B$ respectively. Then $W_1(\mu_A, \mu_B) \leq \sqrt{2n}$.*

**Proof of Claim 2.3.** Observe that any bijection $\phi$ from $A$ to $B$ naturally defines a coupling $q$ of $\mu_A$ and $\mu_B$, defined as

$$q(x, y) = \begin{cases} \frac{1}{|A|} & \text{if } x \in A \text{ and } y = \phi(x), \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $W_1(\mu_A, \mu_B) \leq \mathbb{E}_{x \in A}[\text{dist}(x, \phi(x))]$.

For the other direction note that in the definition of $W_1$ we are looking for the infimum of the linear function $L(q) = \sum_{(x,y) \in A \times B} \text{dist}(x, y) q(x, y)$, where the infimum is taken over the *Birkhoff polytope* of all $n \times n$ doubly stochastic matrices. By the Birkhoff-von Neumann theorem [3, 18] this polytope is the convex hull whose extremal points are precisely the permutation matrices. The optimum is obtained on such an extremal point, and hence there exists a bijection $\phi$ from $A$ to $B$ such that $W_1(\mu_A, \mu_B) = \mathbb{E}[\text{dist}(x, \phi(x))]$. □

**Proof of Claim 2.4.** The proof of the claim follows rather directly from the techniques in *transportation theory* (see [[12], Section 3.4]). Specifically, using Definition 3.4.2 and combining Proposition 3.4.1, equation 3.4.42, and Proposition 3.4.3, where $\mathcal{X} = \{0, 1\}$, and $\mu$ is the uniform distribution on $\mathcal{X}$ we have the following theorem. □

**Theorem 2.5.** *Let $\nu$ be an arbitrary distribution on the discrete hypercube $\mathcal{H}_n$, and let $\mu_n$ be the uniform distribution on $\mathcal{H}_n$. Then*

$$W_1(\nu, \mu_n) \leq \sqrt{\frac{1}{2}n \cdot D(\nu \,||\, \mu_n)},$$

*where $D(\nu \,||\, \mu)$ is the Kullback-Leibler divergence defined as $D(\nu \,||\, \mu) = \sum_x \nu(x) \log\left(\frac{\nu(x)}{\mu(x)}\right)$.*

In particular, by letting $\nu = \mu_A$ be the uniform distribution over the set $A$ of cardinality $2^{n-1}$, we have $D(\mu_A \,||\, \mu_n) = \sum_{x \in A} \mu_A(x) \log\left(\frac{\mu_A(x)}{\mu_n(x)}\right) = \sum_{x \in A} \frac{1}{|A|} \log(2) = 1$, and hence $W_1(\mu_n, \mu_A) \leq \sqrt{\frac{1}{2}n \cdot D(\mu_n \,||\, \nu)} = \sqrt{n/2}$. Analogously, we have $W_1(\mu_n, \mu_B) \leq \sqrt{n/2}$. Therefore, by the triangle inequality, we conclude that $W_1(\mu_A, \mu_B) \leq W_1(\mu_A, \mu_n) + W_1(\mu_n, \mu_B) \leq \sqrt{2n}$, as required.

This completes the proof of Theorem 1.3.

## 3. Average stretch for recursive majority of 3's

In this section we prove Theorem 1.6, showing a mapping from $\mathcal{H}_n$ to $A_{\text{rec-maj}_k}$ with constant average stretch. The key step in the proof is the following lemma.

**Lemma 3.1.** *Let $k$ be a positive integer, and let $n = 3^k$. There exists $f_k : \mathcal{H}_n \to A_{\text{rec-maj}_k}$ satisfying the following properties.*

1. *$f_k(x) = x$ for all $x \in A_{\text{rec-maj}_k}$.*
2. *For each $x \in A_{\text{rec-maj}_k}$ there is a unique $z \in Z_{\text{rec-maj}_k} := \mathcal{H}_n \setminus A_{\text{rec-maj}_k}$ such that $f_k(z) = x$.*
3. *For every $i \in [n]$ we have $\mathbb{E}_{x \in \mathcal{H}_n}\left[\text{dist}\left(f_k(x), f_k(x + e_i)\right)\right] \leq 10$.*

We postpone the proof of Lemma 3.1 for now, and show how it implies Theorem 1.6.

**Proof of Theorem 1.6.** Let $f_k$ be the mapping from Lemma 3.1. Define $\psi_0, \psi_1 : \mathcal{H}_{n-1} \to A_{\text{rec-maj}_k}$ as $\psi_b(x) = f_k(x \circ b)$, where $x \circ b \in \mathcal{H}_n$ is the string obtained from $x$ by appending to it $b$ as the $n$'th coordinate.

The mappings $\psi_0, \psi_1$ naturally induce a bipartite graph $G = (V, E)$, where $V = \mathcal{H}_{n-1} \cup A_{\text{rec-maj}_k}$ and $E = \{(x, \psi_b(x)) : x \in \mathcal{H}_{n-1}, b \in \{0, 1\}\}$, possibly, containing parallel edges. Note that by the first two items of Lemma 3.1 the graph $G$ is 2-regular. Indeed, for each $x \in \mathcal{H}_n$ the neighbours of $x$ are $N(x) = \{\psi_0(x) = f_k(x \circ 0), \psi_1(x) = f_k(x \circ 1)\}$, and for each $y \in A_{\text{rec-maj}_k}$ there is a unique $x \in A_{\text{rec-maj}_k}$ and a unique $z \in Z_{\text{rec-maj}_k}$ such that $f_k(x) = f_k(z) = 1$, and hence $N(y) = \{x_{[1,\ldots,n-1]}, z_{[1,\ldots,n-1]}\}$.

Since the bipartite graph $G$ is 2-regular, it has a perfect matching. Let $\phi$ be the bijection from $\mathcal{H}_{n-1}$ to $A_{\text{rec-maj}_k}$ induced by a perfect matching in $G$, and for each $x \in \mathcal{H}_n$ let $b_x \in \mathcal{H}_n$ be such that $\phi(x) = \psi_{b_x}(x)$. We claim that $\mathsf{avgStretch}(\phi) = O(1)$. Let $x \sim x'$ be uniformly random vertices in $\mathcal{H}_{n-1}$ that differ in exactly one coordinate, and let $r \in \{0, 1\}$ be uniformly random. Then

$$
\begin{aligned}
\mathbb{E}\left[\mathsf{dist}\left(\phi(x), \phi\left(x'\right)\right)\right] &= \mathbb{E}\left[\mathsf{dist}\left(f_k\left(x \circ b_x\right), f_k\left(x' \circ b_{x'}\right)\right)\right] \\
&\leq \mathbb{E}\left[\mathsf{dist}\left(f_k\left(x \circ b_x\right), f_k\left(x \circ r\right)\right)\right] + \mathbb{E}\left[\mathsf{dist}\left(f_k\left(x \circ r\right), f_k\left(x' \circ r\right)\right)\right] \\
&\quad + \mathbb{E}\left[\mathsf{dist}\left(f_k\left(x' \circ r\right), f_k\left(x' \circ b_{x'}\right)\right)\right].
\end{aligned}
$$

For the first term, since $r$ is equal to $b_x$ with probability $1/2$ by Lemma 3.1 Item 3 we get that $\mathbb{E}\left[\mathsf{dist}\left(f_k\left(x \circ b_x\right), f_k(x \circ r)\right)\right] \leq 5$. Analogously the third term is bounded by 5. In the second term we consider the expected distance between $f(\cdot)$ applied on inputs that differ in a random coordinate $i \in [n-1]$, which is at most 10, again, by Lemma 3.1 Item 3. Therefore, $\mathbb{E}\left[\mathsf{dist}\left(\phi(x), \phi\left(x'\right)\right)\right] \leq 20$. $\qquad\square$

We return to the proof of Lemma 3.1.

**Proof of Lemma 3.1.** Define $f_k : \mathcal{H}_n \to A_{\text{rec-maj}_k}$ by induction on $k$. For $k = 1$ define $f_1$ as

$$
\begin{aligned}
000 &\mapsto 110 \\
100 &\mapsto 101 \\
010 &\mapsto 011 \\
001 &\mapsto 111 \\
x &\mapsto x \quad \text{for all } x \in \{110, 101, 011, 111\}.
\end{aligned}
$$

That is, $f_1$ acts as the identity map for all $x \in A_{\text{rec-maj}_1}$, and maps all inputs in $Z_{\text{rec-maj}_1}$ to $A_{\text{rec-maj}_1}$ in a one-to-one way. Note that $f_1$ is a non-decreasing mapping, that is, $(f_1(x))_i \geq x_i$ for all $x \in \mathcal{H}_3$ and $i \in [3]$.

For $k > 1$ define $f_k$ recursively using $f_{k-1}$ as follows. For each $r \in [3]$, let $T_r = [(r-1) \cdot 3^{k-1} + 1, \ldots, r \cdot 3^{k-1}]$ be the $r$'th third of the interval $[3^k]$. For $x \in \mathcal{H}_{3^k}$, write $x = x^{(1)} \circ x^{(2)} \circ x^{(3)}$, where $x^{(r)} = x_{T_r} \in \mathcal{H}_{3^{k-1}}$ is the $r$'th third of $x$. Let $y = (y_1, y_2, y_3) \in \{0, 1\}^3$ be defined as $y_r = \text{rec-maj}_{k-1}\left(x^{(r)}\right)$, and let $w = (w_1, w_2, w_3) = f_1(y) \in \{0, 1\}^3$. Define

$$
f_{k-1}^{(r)}\left(x^{(r)}\right) = \begin{cases} f_{k-1}\left(x^{(r)}\right) & \text{if } w_r \neq y_r, \\ x^{(r)} & \text{otherwise.} \end{cases}
$$

Finally, the mapping $f_k$ is defined as

$$
f_k(x) = f_{k-1}^{(1)}\left(x^{(1)}\right) \circ f_{k-1}^{(2)}\left(x^{(2)}\right) \circ f_{k-1}^{(3)}\left(x^{(3)}\right).
$$

That is, if $\text{rec-maj}_k(x) = 1$ then $w = y$, and hence $f_k(x) = x$, and otherwise, $f_{k-1}^{(r)}\left(x^{(r)}\right) \neq x^{(r)}$ for all $r \in [3]$ where $y_r = 0$ and $w_r = 1$.

Next we prove that $f_k$ satisfies the properties stated in Lemma 3.1.

1. It is clear from the definition that if $\text{rec-maj}_k(x) = 1$, then $w = y$, and hence $f_k(x) = x$.

2. Next, we prove by induction on $k$ that the restriction of $f_k$ to $Z_{\text{rec-maj}_k}$ induces a bijection. For $k = 1$ the statement clearly holds. For $k > 2$ suppose that the restriction of $f_{k-1}$ to $Z_{\text{rec-maj}_{k-1}}$ induces a bijection. We show that for every $x \in A_{\text{rec-maj}_k}$ the mapping $f_k$ has a preimage of $x$ in $Z_{\text{rec-maj}_k}$. Write $x = x^{(1)} \circ x^{(2)} \circ x^{(3)}$, where $x^{(r)} = x_{T_r} \in \mathcal{H}_{3^{k-1}}$ is the $r$'th third of $x$. Let $w = (w_1, w_2, w_3)$ be defined as $w_r = \text{rec-maj}_k[k-1](x^{(r)})$. Since $x \in A_{\text{rec-maj}_k}$ it follows that $w \in \{110, 101, 011, 111\}$. Let $y = (y_1, y_2, y_3) \in Z_{\text{rec-maj}_1}$ such that $f_1(y) = w$. For each $r \in [3]$ such that $w_r = 1$ and $y_r = 0$ it must be the case that $x^{(r)} \in A_{\text{rec-maj}_{k-1}}$, and hence, by the induction hypothesis, there is some $z^{(r)} \in Z_{\text{rec-maj}_{k-1}}$ such that $f_{k-1}(z^{(r)}) = x^{(r)}$. For each $r \in [3]$ such that $y_r = w_r$ define $z^{(r)} = x^{(r)}$. Since $y = (y_1, y_2, y_3) \in Z_{\text{rec-maj}_1}$, it follows that $z = z^{(1)} \circ z^{(2)} \circ z^{(3)} \in Z_{\text{rec-maj}_k}$. It is immediate by the construction that, indeed, $f_k(z) = x$.

3. Fix $i \in [3^k]$. In order to prove $\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i))] = O(1)$ consider the following events.

$$E_1 = \{\text{rec-maj}_k(x) = 1 = \text{rec-maj}_k(x + e_i)\},$$
$$E_2 = \{\text{rec-maj}_k(x) = 0, \text{rec-maj}_k(x + e_i) = 1\},$$
$$E_3 = \{\text{rec-maj}_k(x) = 1, \text{rec-maj}_k(x + e_i) = 0\},$$
$$E_4 = \{\text{rec-maj}_k(x) = 0 = \text{rec-maj}_k(x + e_i)\}.$$

Then $\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i))] = \sum_{j=1,2,3,4} \mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) | E_j] \cdot \mathbb{P}[E_j]$. The following three claims prove an upper bound on $\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i))]$.

**Claim 3.2.** $\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) | E_1] = 1.$

**Claim 3.3.** $\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) | E_2] \leq 2 \cdot 1.5^k.$

**Claim 3.4.** $\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) | E_4] \cdot \mathbb{P}[E_4] \leq 8.$

By symmetry, it is clear that $\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) | E_2] = \mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) | E_3]$. Note also that $\mathbb{P}[E_1] < 0.5$ and $\mathbb{P}[E_2 \cup E_3] = 2^{-k}$.[3] Therefore, the claims above imply that

$$\mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i))] = \sum_{j=1,2,3,4} \mathbb{E}[\text{dist}(f_k(x), f_k(x + e_i)) | E_i] \cdot \mathbb{P}[E_i] \leq 1 \cdot 0.5$$
$$+ 2 \cdot 1.5^k \cdot 2^{-k} + 8 \leq 10,$$

which completes the proof of Lemma 3.1.

Next we prove the above claims.

**Proof of Claim 3.2.** If $E_1$ holds then $\text{dist}(f_k(x), f_k(x + e_i)) = \text{dist}(x, x + e_i) = 1.$ ☐

**Proof of Claim 3.3.** We prove first that

$$\mathbb{E}[\text{dist}(x, f_k(x)) | \text{rec-maj}_k(x) = 0] = 1.5^k. \tag{1}$$

---

[3] Indeed, note that $\mathbb{P}[E_2 \cup E_3] = \mathbb{P}[\text{rec-maj}_k(x) \neq \text{rec-maj}_k(x + e_i)]$, and suppose for concreteness that $i = 1$. We claim that $\mathbb{P}[\text{rec-maj}_k(x) \neq \text{rec-maj}_k(x + e_1)] = 2^{-k}$, which can be seen by induction on $k$ using the recurrence $\mathbb{P}[\text{rec-maj}_k(x) \neq \text{rec-maj}_k(x + e_1)] = \mathbb{P}[\text{rec-maj}_{k-1}(x^{(2)}) \neq \text{rec-maj}_{k-1}(x^{(3)})] \cdot \mathbb{P}[\text{rec-maj}_{k-1}(x^{(1)}) \neq \text{rec-maj}_{k-1}(x^{(1)} + e_1)]$ $= (1/2) \cdot \mathbb{P}[\text{rec-maj}_{k-1}(x^{(1)}) \neq \text{rec-maj}_{k-1}(x^{(1)} + e_1)] = (1/2) \cdot 2^{-(k-1)} = 2^{-k}$.

The proof is by induction on $k$. For $k=1$ we have $\mathbb{E}\left[\text{dist}\left(x, f_1(x)\right) | \text{rec-maj}_k(x) = 0\right] = 1.5$ as there are two inputs $x \in Z_{\text{rec-maj}_k}$ with $\text{dist}\left(x, f_1(x)\right) = 1$ and two $x$'s in $Z_{\text{rec-maj}_k}$ with $\text{dist}\left(x, f_1(x)\right) = 2$. For $k > 1$ suppose that $\mathbb{E}\left[\text{dist}(x, f_{k-1}(x)) | \text{rec-maj}_{k-1}(x)\right] = 1.5^{k-1}$. Write each $x \in \mathcal{H}_{3^k}$ as $x = x^{(1)} \circ x^{(2)} \circ x^{(3)}$, where $x^{(r)} = x_{T_r} \in \mathcal{H}_{3^{k-1}}$ is the $r$'th third of $x$, and let $y = (y_1, y_2, y_3)$ be defined as $y_r = \text{rec-maj}_{k-1}\left(x^{(r)}\right)$. Since $\mathbb{E}_{x \in H_{3^{k-1}}}\left[\text{rec-maj}_{k-1}(x)\right] = 0.5$, it follows that for a random $z \in Z_{\text{rec-maj}_k}$ each $y \in \{000, 100, 010, 001\}$ happens with the same probability $1/4$, and hence, using the induction hypothesis we get

$$\mathbb{E}\left[\text{dist}\left(x, f_k(x)\right) | \text{rec-maj}_k(x) = 0\right] = \mathbb{P}\left[y \in \{100, 010\} | \text{rec-maj}_k(x) = 0\right] \times 1.5^{k-1}$$
$$+ \mathbb{P}\left[y \in \{000, 001\} | \text{rec-maj}_k(x) = 0\right] \times 2 \cdot 1.5^{k-1}$$
$$= 1.5^k,$$

which proves equation (1).

Next we prove that[4]

$$\mathbb{E}\left[\text{dist}\left(x, f_k(x)\right) | E_2\right] \le \sum_{j=0}^{k-1} 1.5^j = 2 \cdot \left(1.5^k - 1\right). \tag{2}$$

Note that equation (2) proves Claim 3.3. Indeed, if $E_2$ holds then using the triangle inequality we have $\text{dist}\left(f_k(x), f_k(x + e_i)\right) \le \text{dist}(f_k(x), x) + \text{dist}(x, x + e_i) + \text{dist}(x + e_i, f_k(x + e_i)) = \text{dist}(f_k(x), x) + 1$, and hence

$$\mathbb{E}[\text{dist}(f_k(x), x) | E_2] + 1 \le 2 \cdot (1.5^k - 1) + 1 < 2 \cdot 1.5^k,$$

as required.

We prove equation (2) by induction on $k$. For $k = 1$ equation (2) clearly holds. For the induction step let $k > 1$. As in the definition of $f_k$ write each $x \in \mathcal{H}_{3^k}$ as $x = x^{(1)} \circ x^{(2)} \circ x^{(3)}$, where $x^{(r)} = x_{T_r}$ is the $r$'th third of $x$, and let $y = (y_1, y_2, y_3)$ be defined as $y_r = \text{rec-maj}_{k-1}\left(x^{(r)}\right)$.

Let us suppose for concreteness that $i \in T_1$. (The cases of $i \in T_2$ and $i \in T_3$ are handled similarly.) Note that if $\text{rec-maj}_k(x) = 0$, $\text{rec-maj}_k(x + e_i) = 1$, and $i \in T_1$, then $y \in \{010, 001\}$. We consider each case separately.

1. Suppose that $y = 010$. Then $w = f(y) = 011$, and hence $f(x)$ differs from $x$ only in $T_3$. Taking the expectation over $x$ such that $\text{rec-maj}_k(x) = 0$ and $\text{rec-maj}_k(x + e_i) = 1$ by equation (1) we get $\mathbb{E}\left[\text{dist}(x, f(x)) | E_2, y = 010\right] = \mathbb{E}\left[\text{dist}\left(f_{k-1}(x^{(3)}), x^{(3)}\right) | \text{rec-maj}_{k-1}\left(x^{(3)}\right) = 0\right] = 1.5^{k-1}$.

2. If $y = 001$, then $w = f_1(y) = 111$, and $f(x)$ differs from $x$ only in $T_1 \cup T_2$. Then

$$\mathbb{E}\left[\text{dist}(x, f(x)) | E_2, y = 001\right] = \mathbb{E}\left[\text{dist}\left(f_{k-1}(x^{(1)}), x^{(1)}\right) | E_2, y = 001\right]$$
$$+ \mathbb{E}\left[\text{dist}\left(f_{k-1}(x^{(2)}), x^{(2)}\right) | E_2, y = 001\right].$$

Denoting by $E_2'$ the event that $\text{rec-maj}_{k-1}\left(x^{(1)}\right) = 0$, $\text{rec-maj}_{k-1}(x^{(1)} + e_i) = 1$ (i.e., the analogue of the event $E_2$ applied on $\text{rec-maj}_{k-1}$), we note that

$$\mathbb{E}\left[\text{dist}\left(f_{k-1}(x^{(1)}), x^{(1)}\right) | E_2, y = 001\right] = \mathbb{E}\left[\text{dist}\left(f_{k-1}(x^{(1)}), x^{(1)}\right) | E_2'\right],$$

---

[4]Note that equation (2) can be thought of equation (1) conditioned on the event $\text{rec-maj}_k(x + e_i) = 1$, which happens with probability only $2^{-k}$. A naive application of Markov's inequality would only say that $\mathbb{E}[\text{dist}(x, f_k(x)) | E_2] \le 1.5^k \cdot 2^k$, which would not suffice for us. Equation (2) says that the expected distance is comparable to $1.5^k$ even when conditioning on this small event.

which is upper bounded by $\sum_{j=0}^{k-2} 1.5^j$ using the induction hypothesis. For the second term we have

$$\mathbb{E}\left[\mathsf{dist}\big(f_{k-1}\big(x^{(2)}\big), x^{(2)}\big)|E_2, y=001\right] = \mathbb{E}\left[\mathsf{dist}\big(f_{k-1}\big(x^{(2)}\big), x^{(2)}\big)|\mathsf{rec\text{-}maj}_{k-1}\big(x^{(2)}\big)=0\right],$$

which is at most $1.5^{k-1}$ using equation (1). Therefore, for $y=001$ we have

$$\mathbb{E}[\mathsf{dist}(x, f(x))|E_2, y=001] \le \sum_{j=0}^{k-2} 1.5^j + 1.5^{k-1}.$$

Using the two cases for $y$ we get

$$\begin{aligned}
\mathbb{E}\big[\mathsf{dist}\big(x, f_k(x)\big)|E_2\big] &= \mathbb{E}\big[\mathsf{dist}\big(x, f_k(x)\big)|E_2, y=010\big] \cdot \mathbb{P}\big[y=010|E_2\big] \\
&\quad + \mathbb{E}\big[\mathsf{dist}\big(x, f_k(x)\big)|E_2, y=001\big] \cdot \mathbb{P}\big[y=001|E_2\big] \\
&\le \sum_{j=0}^{k-1} 1.5^j.
\end{aligned}$$

This proves equation (2) for the case where $i \in T_1$. The other two cases are handled similarly. This completes the proof of Claim 3.3. $\qquad\square$

**Proof of Claim 3.4.** For a coordinate $i \in [n]$ and for $0 \le j \le k$ let $r = r_i(j) \in \mathbb{N}$ be such that $i \in [(r-1) \cdot 3^j + 1, \ldots, r \cdot 3^j]$, and denote the corresponding interval by $T_i(j) = [(r-1) \cdot 3^j + 1, \ldots, r \cdot 3^j]$.[5] These are the coordinates used in the recursive definition of $\mathsf{rec\text{-}maj}_k$ by the instance of $\mathsf{rec\text{-}maj}_j$ that depends on the $i$'th coordinate.

For $x \in \mathcal{H}_n$ and $x' = x + e_i$, define $v(x)$ as

$$v(x) = \begin{cases} \min\left\{j \in [k] : \mathsf{rec\text{-}maj}_j\big(x_{T_i(j)}\big) = \mathsf{rec\text{-}maj}_j\big(x'_{T_i(j)}\big)\right\} & \text{if } \mathsf{rec\text{-}maj}_k(x) = \mathsf{rec\text{-}maj}_k(x'), \\ k+1 & \text{if } \mathsf{rec\text{-}maj}_k(x) \ne \mathsf{rec\text{-}maj}_k(x'). \end{cases}$$

That is, in the ternary tree defined by the computation of $\mathsf{rec\text{-}maj}_k$, $v(x)$ is the lowest $j$ on the path from the $i$'th coordinate to the root where the computation of $x$ is equal to the computation of $x + e_i$. Note that if $x$ is chosen uniformly from $\mathcal{H}_n$, then

$$\mathbb{P}[v=j] = \begin{cases} 2^{-j} & \text{if } j \in [k], \\ 2^{-k} & \text{if } j = k+1. \end{cases} \tag{3}$$

Below we show that by conditioning on $E_4$ and on the value of $v$ we get

$$\mathbb{E}\left[\mathsf{dist}\big(f_k(x), f_k(x + e_i)\big)|E_4, v=j\right] \le 4 \cdot 1.5^j. \tag{4}$$

Indeed, suppose that $E_4$ holds. Assume without loss of generality that $x_i = 0$, and let $x' = x + e_i$. Note that $f_k(x)$ and $f_k(x')$ differ only on the coordinates in the interval $T_i(v)$. Let $w = x_{T_i(v)}$, and define $y = (y_1, y_2, y_3) \in \{0,1\}^3$ as $y_r = \mathsf{rec\text{-}maj}_{v-1}(w^{(r)})$ for each $r \in [3]$, where $w^{(r)}$ is the $r$'th third of $w$. Similarly, let $w' = x'_{T_i(v)}$, and let $y' = (y'_1, y'_2, y'_3) \in \{0,1\}^3$ be defined as $y'_r = \mathsf{rec\text{-}maj}_{v-1}(w'^{(r)})$ for each $r \in [3]$. This implies that

$$\mathbb{E}\left[\mathsf{dist}\big(f_k(x), f_k(x + e_i)\big)|E_4\right] = \mathbb{E}\left[\mathsf{dist}\big(f_v(w), f_v(w')\big)|E_4\right].$$

Furthermore, if $\mathsf{rec\text{-}maj}_v\big(x_{T_i(v)}\big) = 1$ $\big($and $\mathsf{rec\text{-}maj}_v\big(x'_{T_i(v)}\big) = 1\big)$, then $f_k(x)_{T_i(v)} = x_{T_i(v)}$, and thus $\mathsf{dist}\big(f_k(x), f_k(x')\big) = 1$.

---

[5]For example, for $j = 0$ we have $T_i(0) = \{i\}$. For $j = 1$ if $i = 1 \pmod 3$ then $T_i(1) = [i, i+1, i+2]$. For $j = k-1$ the interval $T_i(k-1)$ is one of the intervals $T_1, T_2, T_3$. For $j = k$ we have $T_i(k) = [1, \ldots, 3^k]$.

Next we consider the case of $\mathsf{rec\text{-}maj}_k\big(x_{T_i(v)}\big) = 0$ $\big($and $\mathsf{rec\text{-}maj}_k\big(x'_{T_i(v)}\big) = 0\big)$. Since $x_i = 0$ and $x' = x + e_i$, it must be that $y = 000$ and $y'$ is a unit vector. Suppose first that $y' = 100$, that is, the coordinate $i$ belongs to the first third of $T_i(v)$. Write $w = w^{(1)} \circ w^{(2)} \circ w^{(3)}$, where each $w^{(r)}$ is one third of $w$. Analogously, write $w' = w'^{(1)} \circ w'^{(2)} \circ w'^{(3)}$, where each $w'^{(r)}$ one third of $w'$. Then, since $w' = w + e_i$ we have

$$
\begin{aligned}
\mathbb{E}\left[\mathsf{dist}\left(f_j(w), w\right) | E_4, v = j\right] &= \mathbb{E}\Big[\mathsf{dist}\left(f_{j-1}(w^{(1)}), w^{(1)}\right) | \mathsf{rec\text{-}maj}_{j-1}\big(w^{(1)}\big) \\
&= 0, \mathsf{rec\text{-}maj}_{j-1}\big(w^{(1)} + e_i\big) = 1\Big] \\
&\quad + \mathbb{E}\Big[\mathsf{dist}\left(f_{j-1}(w^{(2)}), w^{(2)}\right) | \mathsf{rec\text{-}maj}_{j-1}\big(w^{(2)}\big) = 0\Big] \\
&\leq 2 \cdot \left(1.5^{j-1} - 1\right) + 1.5^{j-1} = 3 \cdot 1.5^{j-1} - 2,
\end{aligned}
$$

where the last inequality is by equation (1) and equation (2). Similarly,

$$
\mathbb{E}\left[\mathsf{dist}\left(f_j(w'), w'\right) | E_4, v = j\right] = \mathbb{E}\Big[\mathsf{dist}\left(f_{j-1}(w'^{(3)}), w'^{(3)}\right) | \mathsf{rec\text{-}maj}_{j-1}\big(w'^{(3)}\big) = 0\Big] \leq 1.5^{j-1},
$$

where the last inequality is by equation (1). Therefore,

$$
\mathbb{E}\left[\mathsf{dist}\left(f_v(w), f_v(w')\right) | E_4, v = j\right] < 4 \cdot 1.5^{j-1}.
$$

The cases of $y = 010$ and $001$ are handled similarly, and it is straightforward to verify that in these cases we also get the bound of $4 \cdot 1.5^{j-1}$.

By combining equation (3) with equation (4) it follows that

$$
\begin{aligned}
\mathbb{E}[\mathsf{dist}\left(f_k(x), f_k(x + e_i)\right) | E_4] \cdot \mathbb{P}[E_4] &= \sum_{j=1}^{k} \mathbb{E}[\mathsf{dist}\left(f_k(x), f_k(x + e_i)\right) | E_4, v = j] \cdot \mathbb{P}[v = j | E_4] \cdot \mathbb{P}[E_4] \\
&\leq \sum_{j=1}^{k} 4 \cdot 1.5^{j-1} \cdot \mathbb{P}[v = j] \\
&\leq 4 \cdot \sum_{j=1}^{k} 1.5^{j-1} \cdot 2^{-j} \leq 8.
\end{aligned}
$$

This completes the proof of Claim 3.4. $\qquad\square$

## 4. Average stretch for tribes

In this section we prove Theorem 1.8, showing a mapping from $\mathcal{H}_n$ to $A^*_{\mathsf{tribes}}$ with $O(\log(n))$ average stretch. Let $\mu^1_{\mathsf{tribes}}$ be the uniform distribution on $A_{\mathsf{tribes}}$, and let $\mu^0_{\mathsf{tribes}}$ be the uniform distribution on $Z_{\mathsf{tribes}} = \mathcal{H}_n \setminus A_{\mathsf{tribes}}$. The proof consists of the following two claims.

**Claim 4.1.** *For $\mu^1_{\mathsf{tribes}}$ and $\mu^0_{\mathsf{tribes}}$ as above it holds that*

$$
W_1\left(\mu^0_{\mathsf{tribes}}, \mu^1_{\mathsf{tribes}}\right) = O(\log(n)).
$$

Next, let $A^*_{\mathsf{tribes}} \subseteq \mathcal{H}_n$ be an arbitrary superset of $A_{\mathsf{tribes}}$ of density $1/2$, and let $\mu^*_{\mathsf{tribes}}$ be the uniform distribution on $A^*_{\mathsf{tribes}}$.

**Claim 4.2.** *Consider $\mathcal{H}_{n-1}$ as $\{x \in \mathcal{H}_n : x_n = 0\}$, and let $\mu_{n-1}$ be the uniform measure on $\mathcal{H}_{n-1}$. Then,*

$$
W_1\left(\mu_{n-1}, \mu^*_{\mathsf{tribes}}\right) \leq W_1\left(\mu^0_{\mathsf{tribes}}, \mu^1_{\mathsf{tribes}}\right) + O(\log(n)).
$$

By combining Claim 4.1 and Claim 4.2 we get that the average transportation distance between $\mathcal{H}_{n-1}$ and $A^*_{\text{tribes}}$ is $W_1\left(\mu_{n-1}, \mu^*_{\text{tribes}}\right) = O(\log(n))$. By Claim 2.3 it follows that there exists $\phi_{\text{tribes}} : \mathcal{H}_{n-1} \to A^*_{\text{tribes}}$ such that $\mathbb{E}[\text{dist}(x, \phi_{\text{tribes}}(x))] = O(\log(n))$, and using Proposition 1.4 we conclude that $\text{avgStretch}(\phi_{\text{tribes}}) = O(\log(n))$. This completes the proof of Theorem 1.8.

Below we prove Claim 4.1 and Claim 4.2.

**Proof of Claim 4.1.** Let $\mathcal{D} = \mathcal{D}_w$ the uniform distribution over $\{0,1\}^w \setminus \{1^w\}$, let $p = 2^{-w}$, and denote by $\mathcal{L} = \mathcal{L}_{w,s}$ the binomial distribution $\text{Bin}(p,s)$ conditioned on the outcome being positive. That is,

$$\mathbb{P}[\mathcal{L} = \ell] = \frac{\binom{s}{\ell} p^\ell (1-p)^{s-\ell}}{\sum_{j=1}^s \binom{s}{j} p^j (1-p)^{s-j}} \quad \forall \ell \in \{1, \ldots, s\}.$$

Note that $\mu^0_{\text{tribes}}$ is equal to the product distribution $\mathcal{D}^s$. Note also that in order to sample from the distribution $\mu^1_{\text{tribes}}$, we can first sample $\mathcal{L} \in \{1, \ldots, s\}$, then choose $\mathcal{L}$ random tribes that vote unanimously 1, and for the remaining $s - \mathcal{L}$ tribes sample their values in this tribe according to $\mathcal{D}$.

We define a coupling $q_{\text{tribes}}$ between $\mu^0_{\text{tribes}}$ and $\mu^1_{\text{tribes}}$ as follows. First sample $x$ according to $\mu^0_{\text{tribes}}$. Then, sample $\mathcal{L} \in \{1, \ldots, s\}$, choose $\mathcal{L}$ tribes $T \subseteq [s]$ uniformly, and let $S = \{(t-1)w + j : t \in T, j \in [w]\}$ be all the coordinates participating in all tribes in $T$. Define $y \in \mathcal{H}_n$ as $y_i = 1$ for all $i \in S$, and $y_i = x_i$ for all $i \in [n] \setminus S$. It is clear that $y$ is distributed according to $\mu^1_{\text{tribes}}$, and hence $q_{\text{tribes}}$ is indeed a coupling between $\mu^0_{\text{tribes}}$ and $\mu^1_{\text{tribes}}$.

We next show that $\mathbb{E}_{(x,y) \sim q_{\text{tribes}}}[\text{dist}(x,y)] = O(\log(n))$. We have $\mathbb{E}_{(x,y) \sim q_{\text{tribes}}}[\text{dist}(x,y)] \le \mathbb{E}[\mathcal{L} \cdot w]$, and by the choice of parameters, we have $w \le \log(n)$ and $\mathbb{E}[\mathcal{L}] = \frac{\mathbb{E}[\text{Bin}(2^{-w},s)]}{1 - \mathbb{P}[\text{Bin}(2^{-w},s)=0]} = \frac{s \cdot 2^{-w}}{1 - 2^{-ws}}$. By the choice of $s \le \ln(2) 2^w + O(1)$ it follows that $\mathbb{E}[\mathcal{L}] = O(1)$, and hence

$$W_1\left(\mu^0_{\text{tribes}}, \mu^1_{\text{tribes}}\right) \le \mathbb{E}_{(x,y) \sim q_{\text{tribes}}}[\text{dist}(x,y)] \le \mathbb{E}[\mathcal{L} \cdot w] = O(\log(n)).$$

This completes the proof of Claim 4.1. $\qquad\square$

**Proof of Claim 4.2.** We start by showing that

$$W_1\left(\mu_n, \mu^1_{\text{tribes}}\right) \le W_1\left(\mu^0_{\text{tribes}}, \mu^1_{\text{tribes}}\right), \tag{5}$$

where $\mu_n$ is the uniform measure on $\mathcal{H}_n$. Indeed, let $q_{\text{tribes}}$ be a coupling between $\mu^0_{\text{tribes}}$ and $\mu^1_{\text{tribes}}$. Define a coupling $q_n$ between $\mu_n$ and $\mu^1_{\text{tribes}}$ as

$$q_n(x,y) = \begin{cases} \dfrac{|Z_{\text{tribes}}|}{2^n} \cdot q_{\text{tribes}}(x,y) & \text{if } x \in Z_{\text{tribes}} \text{ and } y \in A_{\text{tribes}}, \\ 1/2^n & \text{if } x = y \in A_{\text{tribes}}, \\ 0 & \text{otherwise.} \end{cases}$$

It is straightforward to verify that $q_n$ is indeed a coupling between $\mu_n$ and $\mu^1_{\text{tribes}}$. Letting $q_{\text{tribes}}$ be a coupling for which $\mathbb{E}_{(x,y) \sim q_{\text{tribes}}}[\text{dist}(x,y)] = W_1\left(\mu^0_{\text{tribes}}, \mu^1_{\text{tribes}}\right)$ we get

$$W_1\left(\mu_n, \mu_{\text{tribes}}^1\right) \leq \sum_{\substack{x \in \mathcal{H}_n \\ y \in A_{\text{tribes}}}} \text{dist}(x,y)q_n(x,y)$$

$$= \sum_{\substack{x \in Z_{\text{tribes}} \\ y \in A_{\text{tribes}}}} \text{dist}(x,y)q_n(x,y) + \sum_{\substack{x \in A_{\text{tribes}} \\ y \in A_{\text{tribes}}}} \text{dist}(x,y)q_n(x,y)$$

$$= \frac{|Z_{\text{tribes}}|}{2^n}\mathbb{E}_{(x,y)\sim q_{\text{tribes}}}[\text{dist}(x,y)] + \sum_{x \in A_{\text{tribes}}} \text{dist}(x,x)q_n(x,x)$$

$$= \frac{|Z_{\text{tribes}}|}{2^n} \cdot W_1\left(\mu_{\text{tribes}}^0, \mu_{\text{tribes}}^1\right) < W_1\left(\mu_{\text{tribes}}^0, \mu_{\text{tribes}}^1\right),$$

which proves equation (5).

Next, we show that

$$W_1\left(\mu_{n-1}, \mu_{\text{tribes}}^1\right) \leq W_1\left(\mu_n, \mu_{\text{tribes}}^1\right) + 1. \tag{6}$$

Indeed, let $q_n$ be a coupling between $\mu_n$ and $\mu_{\text{tribes}}^1$ minimizing $\sum_{(x,y)\in\mathcal{H}_n \times A_{\text{tribes}}} \text{dist}(x,y)q_n(x,y)$. Define a coupling $q_{n-1}$ between $\mu_{n-1}$ and $\mu_{\text{tribes}}^1$ as

$$q_{n-1}(x,y) = q_n(x,y) + q_n(x+e_n, y) \quad \forall x \in \mathcal{H}_{n-1} \text{ and } y \in A_{\text{tribes}}.$$

It is clear that $q_{n-1}$ is a coupling between $\mu_{n-1}$ and $\mu_{\text{tribes}}^1$. Next we prove equation (6).

$$W_1\left(\mu_{n-1}, \mu_{\text{tribes}}^1\right) \leq \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A_{\text{tribes}}}} \text{dist}(x,y)q_{n-1}(x,y)$$

$$= \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A_{\text{tribes}}}} \text{dist}(x,y)q_n(x,y) + \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A_{\text{tribes}}}} \text{dist}(x,y)q_n(x+e_i, y)$$

$$\leq \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A_{\text{tribes}}}} \text{dist}(x,y)q_n(x,y) + \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A_{\text{tribes}}}} (\text{dist}(x+e_i, y)+1)q_n(x+e_i, y)$$

$$= \sum_{\substack{x \in \mathcal{H}_n \\ y \in A_{\text{tribes}}}} \text{dist}(x,y)q_n(x,y) + \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A_{\text{tribes}}}} q_n(x+e_i, y)$$

$$\leq W_1\left(\mu_n, \mu_{\text{tribes}}^1\right) + 1,$$

which proves equation (6).

Next, we show that

$$W_1\left(\mu_{n-1}, \mu_{\text{tribes}}^*\right) \leq W_1\left(\mu_{n-1}, \mu_{\text{tribes}}^1\right) + O(\log(n)). \tag{7}$$

In order to prove equation (7), let $\delta = \frac{1}{2} - \frac{|A_{\text{tribes}}|}{2^n}$. By the discussion in subsection 1.2.2 we have $\delta = O\left(\frac{\log(n)}{n}\right)$. Then $\left|A_{\text{tribes}}^* \setminus A_{\text{tribes}}\right| = \delta \cdot 2^n$. Let $q_{n-1}$ be a coupling between $\mu_{n-1}$ and $\mu_{\text{tribes}}^1$ such that $\mathbb{E}_{(x,y)\sim q_{n-1}}[\text{dist}(x,y)] = W_1\left(\mu_{n-1}, \mu_{\text{tribes}}^1\right)$. Define a coupling $q^*$ between $\mu_{n-1}$ and $\mu_{\text{tribes}}^*$ as

$$q^*(x,y) = \begin{cases} (1-2\delta) \cdot q_{n-1}(x,y) & \text{if } x \in \mathcal{H}_{n-1} \text{ and } y \in A_{\text{tribes}}, \\ 4 \cdot 2^{-2n} & \text{if } x \in \mathcal{H}_{n-1} \text{ and } y \in A_{\text{tribes}}^* \setminus A_{\text{tribes}}. \end{cases}$$

It is straightforward to verify that $q^*$ is a coupling between $\mu_{n-1}$ and $\mu^*_{\text{tribes}}$. Next we prove equation (7).

$$
\begin{aligned}
W_1\left(\mu_{n-1}, \mu^*_{\text{tribes}}\right) &\leq \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A^*_{\text{tribes}}}} \text{dist}(x, y) \cdot q^*(x, y) \\
&= (1 - 2\delta) \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A_{\text{tribes}}}} \text{dist}(x, y) q_{n-1}(x, y) + \sum_{\substack{x \in \mathcal{H}_{n-1} \\ y \in A^*_{\text{tribes}} \setminus A_{\text{tribes}}}} \text{dist}(x, y) \cdot 4 \cdot 2^{-2n} \\
&\leq (1 - 2\delta) \cdot W_1\left(\mu_{n-1}, \mu^1_{\text{tribes}}\right) + 2\delta \cdot \max_{\substack{x \in \mathcal{H}_{n-1} \\ y \in \mathcal{H}_n}} (\text{dist}(x, y)).
\end{aligned}
$$

Equation (7) follows from the fact that $\max(\text{dist}(x, y)) \leq n$ and $\delta = O\left(\frac{\log(n)}{n}\right)$.

By combining equations (5) to (7) we get $W_1\left(\mu_{n-1}, \mu^*_{\text{tribes}}\right) \leq W_1\left(\mu^0_{\text{tribes}}, \mu^1_{\text{tribes}}\right) + O(\log(n))$. $\qquad\square$

## 5. Concluding remarks and open problems

*Uniform upper bound on the average stretch.*   We've shown a uniform upper bound of $O\left(\sqrt{n}\right)$ on the average transportation distance $\mathbb{E}[\text{dist}(x, \phi(x))]$ from $\mathcal{H}_{n-1}$ to any set $A \subseteq \mathcal{H}_n$ of density $1/2$, where $\mathcal{H}_{n-1}$ is treated as $\{x \in \mathcal{H}_n : x_n = 0\}$. This bound is tight up to a multiplicative constant. Indeed, it is not difficult to see that for any bijection $\phi$ from $\mathcal{H}_{n-1}$ to $A_{\text{maj}} = \{x \in \mathcal{H}_n : \sum_i x_i > n/2\}$ (for odd $n$) the average transportation of $\phi$ is $\mathbb{E}[\text{dist}(x, \phi(x))] \geq \Omega\left(\sqrt{n}\right)$.

In contrast, we believe that the upper bound of $O\left(\sqrt{n}\right)$ on the average stretch is not tight, and it should be possible to improve it further.

**Problem 5.1.** *Prove/disprove that for any set $A \subseteq \mathcal{H}_n$ of density $1/2$ there exists a mapping $\phi_A : \mathcal{H}_{n-1} \to A$ with $\text{avgStretch}(\phi) = o\left(\sqrt{n}\right)$.*

*The tribes function.*   Considering our results about the tribes function, we make the following conjecture.

**Conjecture 5.2.** *Let $w$ be a positive integer, and let $s$ be the largest integer such that $1 - (1 - 2^{-w})^s \leq 1/2$. For $n = s \cdot w$ let $\text{tribes} : \mathcal{H}_n \to \{0, 1\}$ be defined as a DNF consisting of $s$ disjoint clauses of width $w$, and let $A_{\text{tribes}} = \{x \in \mathcal{H}_n : \text{tribes}(x) = 1\}$. There exists $A^*_{\text{tribes}} \subseteq \mathcal{H}_n$ a superset of $A_{\text{tribes}}$ of density $\mu_n\left(A^*_{\text{tribes}}\right) = 1/2$ such that $W_1(\mu_{n-1}, \mu^*_{\text{tribes}}) = O(1)$, where $\mu^*_{\text{tribes}}$ is the uniform distribution on $A^*_{\text{tribes}}$.*

As a first step toward the conjecture we proposed the following strengthening of Claim 4.1.

**Problem 5.3.** *Let $\mu^1_{\text{tribes}}$ be the uniform distribution on $A_{\text{tribes}}$, and let $\mu^0_{\text{tribes}}$ be the uniform distribution on $Z_{\text{tribes}} = \mathcal{H}_n \setminus A_{\text{tribes}}$. It is true that $W_1\left(\mu^0_{\text{tribes}}, \mu^1_{\text{tribes}}\right) = O(1)$?*

*A candidate set that requires large average stretch.*   We propose a candidate set $A^*$ for which we hope that any mapping from $\mathcal{H}_{n-1}$ to $A^*$ requires large average stretch. The set is defined as follows. Let $k^* \in [n]$ be the maximal $k$ such that $\binom{n}{\leq k} = \sum_{j=0}^k \binom{n}{j} \leq 2^{n-2}$. Let $B^0_{1/4} = \left\{x \in \mathcal{H}_n : \sum_{i \in [n]} x_i \leq k\right\}$ and $B^1_{1/4} = \left\{x \in \mathcal{H}_n : \sum_{i \in [n]} x_i \geq n - k\right\}$ be two (disjoint) antipodal balls

of radius $k^*$, and let $C \subseteq \mathcal{H}_n \setminus \left( B_{1/4}^0 \cup B_{1/4}^1 \right)$ be an arbitrary set of size $|C| = 2^{n-1} - \left| B_{1/4}^0 \cup B_{1/4}^1 \right|$. Define $A^* = B_{1/4}^0 \cup B_{1/4}^1 \cup C$.

**Conjecture 5.4.** *There is no bijection* $\phi^* : \mathcal{H}_{n-1} \to A^*$ *with* $\mathsf{avgStretch}(\phi^*) = O(1)$.

## Acknowledgements

## References

[1] Angel, O. and Benjamini, I. (2007) A phase transition for the metric distortion of percolation on the hypercube. *Combinatorica* **27**(6) 645–658.

[2] Benjamini, I., Cohen, G. and Shinkar, I. (2014) Bi-Lipschitz bijection between the boolean cube and the Hamming ball, *FOCS*, 81–89.

[3] Birkhoff, G. (1946) Three observations on linear algebra. *Univ. Nac. Tacuman, Rev. Ser. A* **5** 147–151.

[4] Boppana, R. (1997) The average sensitivity of bounded-depth circuits. *Inform. Process. Lett.* **63**(5) 257–261.

[5] Goldreich, O., Goldwasser, S. and Nussboim, A. (2010) On the implementation of huge random objects. *SIAM J. Comput.* **39**(7) 2761–2822.

[6] Graham, R. L. (1988) Isometric embeddings of graphs. *Selected Topics Graph Theory* **3** 133–150.

[7] Håstad, J. (1986) Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM Symposium on Theory of Computing*. ACM. 6–20.

[8] Håstad, J., Leighton, T. and Newman, M. (1987) Reconfiguring a hypercube in the presence of faults. In *Proceedings of the nineteenth annual ACM Symposium on Theory of Computing*, 274–284.

[9] Johnston, T. and Scott, A. (2021) Lipschitz bijections between boolean functions. *Comb. Probab. Comput.* **30**(4) 513–525.

[10] Linial, N. (2002) Finite metric spaces - combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians III*, 573–586.

[11] Linial, N., Mansour, Y. and Nisan, N. (1993) Constant depth circuits, fourier transform, and learnability. *J. ACM* **40**(3) 607–620.

[12] Lovett, S. and Viola, E. (2012) Bounded-depth circuits cannot sample good codes. *Comput. Complex.* **21**(2) 245–266.

[13] O'Donnell, R. (2014) *Analysis of Boolean Functions*. Cambridge: Cambridge University Press.

[14] Raginsky, M. and Sason, I. (2013) Concentration of measure inequalities in information theory, communications, and coding. *Found. Trends Commun. Inform. Theory* **10**(1-2) 1–246.

[15] Rao, S. and Shinkar, I. (2018) On Lipschitz bijections between boolean functions. *Comb. Probab. Comput.* **27**(3) 411–426.

[16] Talagrand, M. (1995) Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. l'Inst. Hautes Études Sci.* **81**(1) 73–205.

[17] Viola, E. (2012) The complexity of distributions. *SIAM J. Comput.* **41**(1) 191–218.

[18] von Neumann, J. (1953) A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contrib. Theory Games II; Ann. Math. Stud.* **28** 5–12.